**LAB ASSIGNMENT 3**

**PYTHON AND DEEP LEARNING**

**Author:**

Harika adivanne

Class id: 02

**Objective:**

The Key objective is to focus on all the regression and prediction models with python machine learning and execute the zero defect code with expected output by using provided datasets.

**Features:**

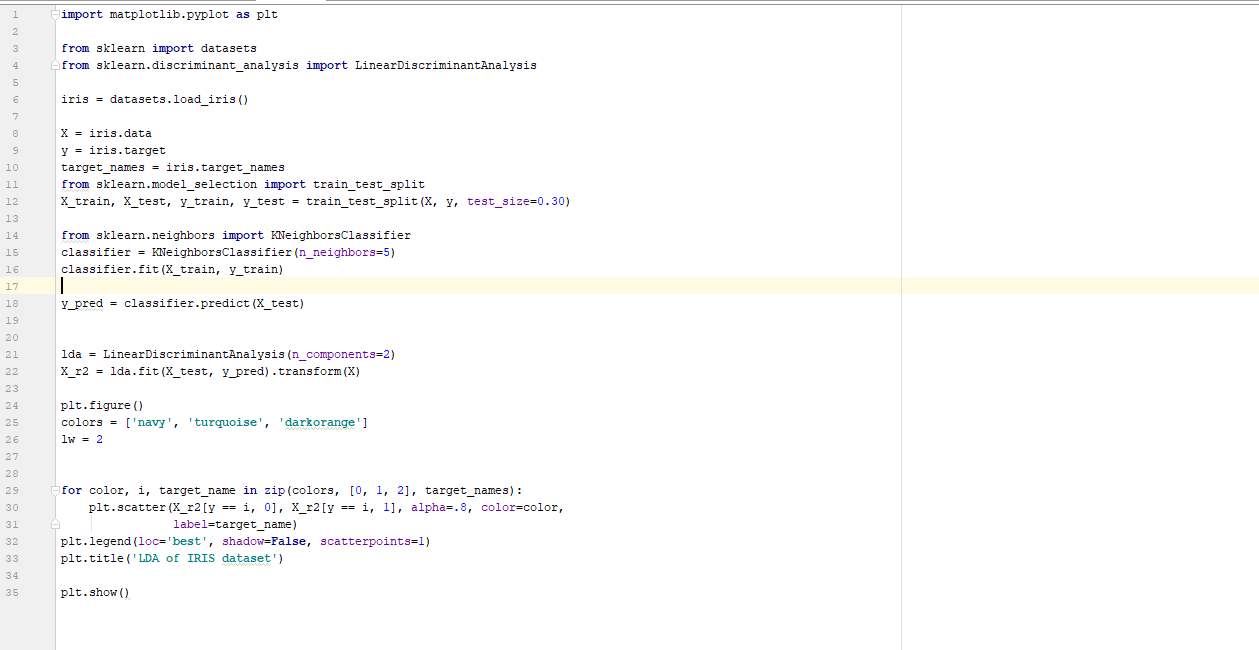
* implemented scikit-learn
* kernel method
* analysis of measured accuracy
* applied lemmatization
* bigrams
* Used Knn algorithm
* Prediction model analysis

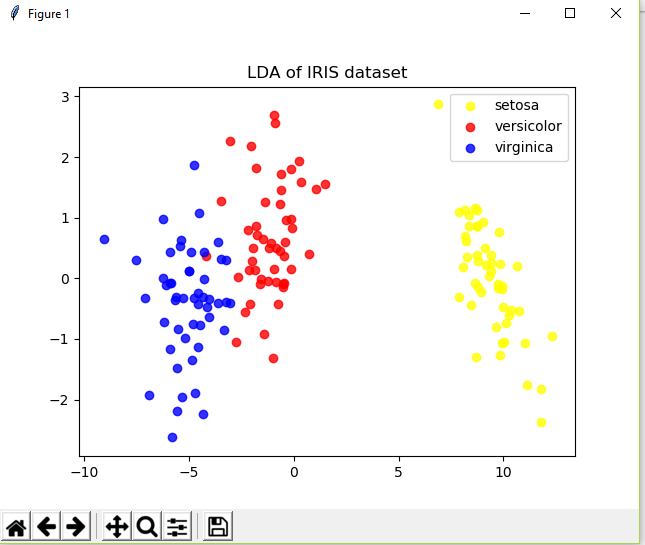
**Configuration:**

* IDE: PycharmCE2017
* LANGUAGE: Python 3

**INPUT/OUTPUT:**

1. **LDA:**

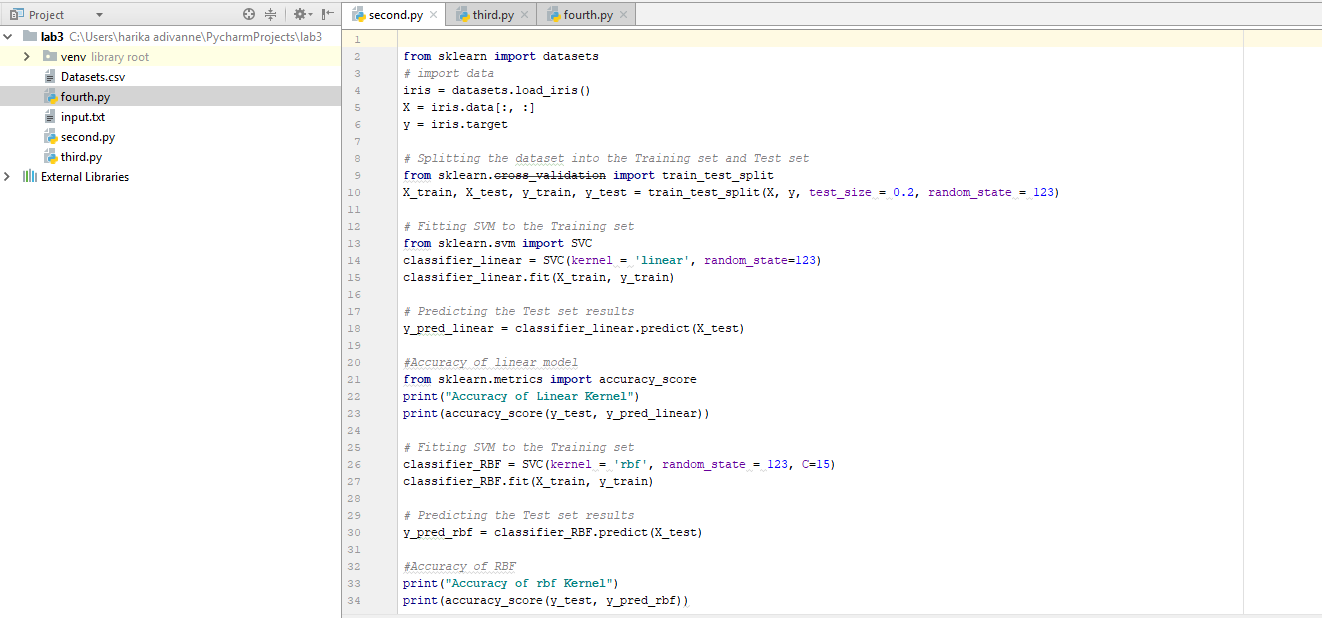
****

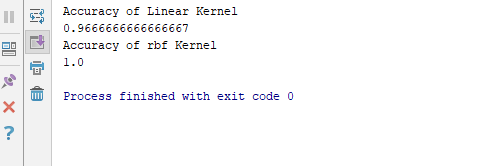
****

**Implementation:**

\*Imported matplot library and linear discriminant analysis. Assigned variables to data and target and also imported kneighbors classifier. Splitted the data into training and testing. Using y\_pred predicted the test data and plotted using three different colors with the title LDA OF IRIS DATASET.

1. **SVM CLASSIFICATION:**

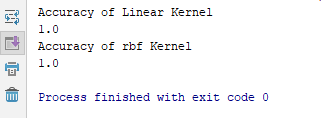
****

****

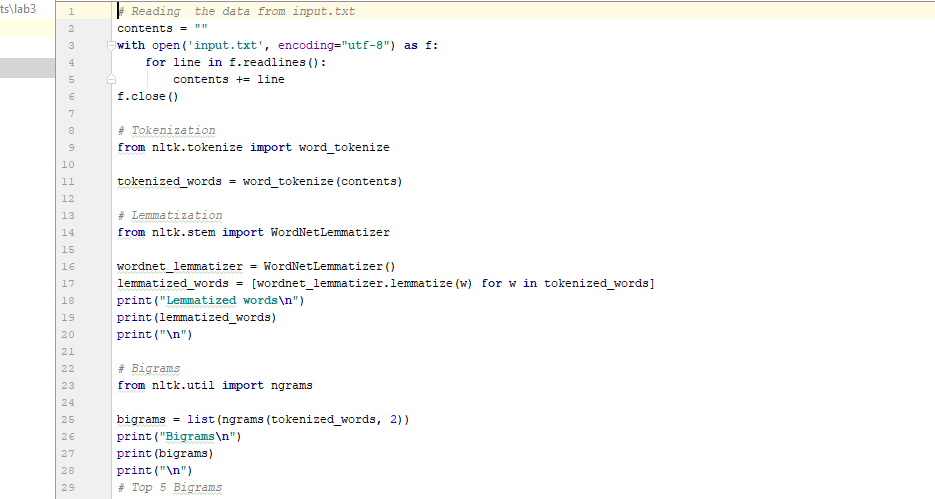
**Implementation:**

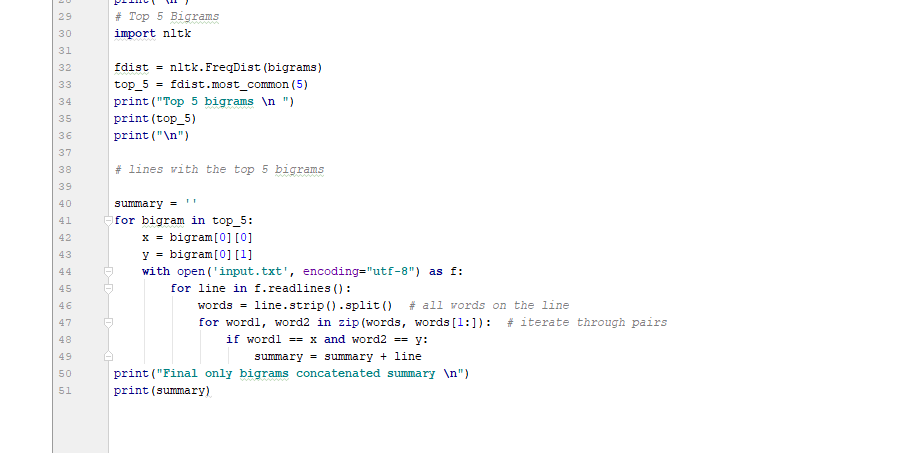
\*Imported data and also installed sklearn package. Data is divided into training and testing sets. Set the SVM data to the training set with the test size of 0.2 and random state=123 and using y\_pred predicted the test results. Printed the accuracy of linear model and then fit SVM into training set and calculated the accuracy of RBF kernel.

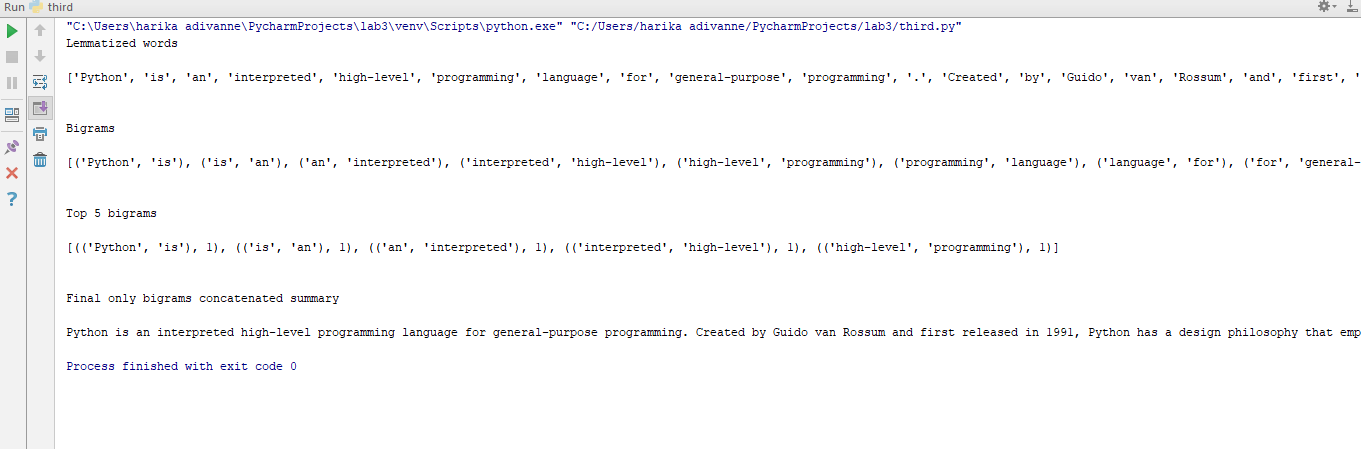
From the above analysis both the kernel and RBF accuracy results similar with a slight difference by changing the c and gamma values we can change the accuracy as equal. I predict RBF kernel is better fit model as class 3 and class 2 of iris data are not linearly separated. Using linear kernel for this data is not ideal.



**3. LEMMATIZATION AND BIGRAM:**

****

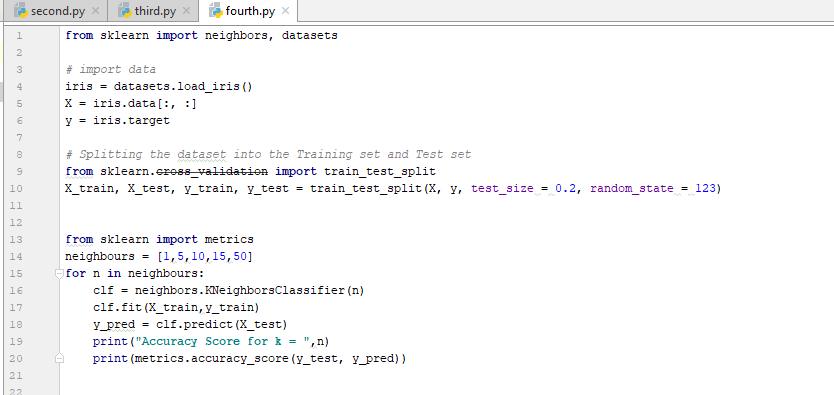
****

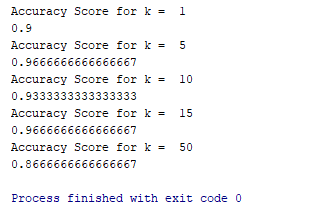
****

**Implementation:**

\*Placed the input text file in IDE and read the data from the text file. Imported nltk package and installed to perform Tokenization, lemmatization and bigrams. Getting the output from tokenization and considering displayed the bigrams out and printed the summary.

1. **K-NEAREST NEIGHBOUR ALGORITHM:**

****

****

**IMPLEMENTATION:**

Imported the iris data and splitted the data into training and testing sets. By using the sklearn package and y\_pred predicted the test results. Finally printed the accuracy for the different values of k ranging from 1 to 50. K with less value will provide better accuracy as k value increases the accuracy level fluctuates as shown in above output.

**Deployment:**

Code has executed in IDE pycharm environment using Python 3 language by creating folder which has all the pycharm files with .py extension. Codes are executed in .py files and the results are shown in console. If it is executed with code 0 then the code has no errors or else it will show the module errors.

**Limitation:**

* Analyzing prediction model
* Understanding the data splitting
* Analysis of accuracy
* Choosing the best suited model

**References:**

<https://www.learnpython.org/>

<https://www.geeksforgeeks.org/python/>